Environ Monit Assess (2020) 192:554
https://doi.org/10.1007/s10661-020-08522-9

Assessment of volunteered geographic information for vegetation mapping

Kellie A. Uyeda · Douglas A. Stow · Casey H. Richart

Received: 25 February 2020 / Accepted: 26 July 2020

Abstract Vegetation mapping requires extensive field data for training and validation. Volunteered geographic information in the form of geotagged photos of identified plants has the potential to serve as a supplemental data source for vegetation mapping projects. In this study, we compare the locations of specific taxa from the iNaturalist platform to locations identified on both a fine-scale vegetation map and high-resolution ortho-imagery in open-canopy shrubland in San Clemente Island, CA. Due to positional uncertainty associated with the iNaturalist observations, as well as the presence-only nature of the data, it was not possible to perform a traditional accuracy assessment. We instead measured the distance between the location recorded by an iNaturalist observer for a given taxon and the closest mapped individual of that taxon. This distance was within 10 m for a majority of the observations (64%). When comparing the iNaturalist location to the closest individual detected through image interpretation, 87% of the observations were within 10 m. The discrepancy in agreement between the vegetation map and imagery is likely due to mapping errors. While iNaturalist data come with important limitations, the platform is an excellent resource for supporting vegetation mapping and other ecological applications.

K. A. Uyeda (✉) · D. A. Stow
Department of Geography, San Diego State University, San Diego, CA, USA
e-mail: kuyeda@sdsu.edu

C. H. Richart
Santa Barbara Botanic Garden, Santa Barbara, CA, USA

Keywords Community-based science · Citizen science · Positional uncertainty · Shrubland · San Clemente Island

Introduction

The last decade has seen a dramatic increase in the capability and usage of websites that facilitate the sharing of volunteered geographic information (VGI) (Goodchild 2007). Datasets available through these websites have potential value for natural resource monitoring and basic ecological research, and have already been demonstrated to be useful in land cover mapping (Fonte et al. 2015), tracking range expansions (Lonhart et al. 2019), detecting elusive species (Richart et al. 2018), phenology tracking (Beaubien and Hall-Beyer 2003), tracking urbanization patterns within conservation zones (Heider et al. 2018), and invasive species monitoring (Taylor et al. 2018), among many other uses. Many other terms are commonly applied to these types of publicly shared datasets, including citizen science, community science, and crowdsourcing (Conrad and Hilchey 2011; Eitzel et al. 2017; See et al. 2016).

VGI platforms vary widely in their goals and the degree to which they are structured. Highly structured platforms typically have a narrowly focused goal, directing users to contribute very specific types of information (Elwood et al. 2013). At the highly structured end, the Degree Confluence project has the goal of collecting photographs of all the latitude and longitude integer degree intersections in the world (http://www.confluence.org). While the purpose of the Degree

Confluence project is not explicitly to support research, it has been used as a reference data source for land cover mapping (Iwao et al. 2006). Other structured platforms focus on observations of specific taxa in defined time periods and locations, such as the Audubon Society's Christmas Bird Count (https://www.audubon.org/conservation/science/christmas-bird-count). An example of a VGI platform with an intermediate level of structure is eBird (ebird.org), which allows users to record lists of birds observed and level of effort expended (length of time) in locations and times defined by the user as well as in localities curated by eBird. On the highly unstructured end are social media websites that allow users to share geotagged photos. These websites can be a valuable source of VGI for ecological research (ElQadi et al. 2017). Even less structured platforms include media reports of unusual phenomena, such as observations of malformed frogs by local schoolchildren (Vandenlangenberg et al. 2003). These unstructured VGI platforms have less oversight and focus than platforms with a highly specific goal for data collection, but even opportunistically gathered data can be tremendously useful for ecological research if the limitations of the dataset are thoughtfully considered (Isaac et al. 2014). Data quality can vary substantially within and across VGI platforms, making it critical to properly assess the quality of the data before analysis is conducted (Senaratne et al. 2017).

The VGI platform iNaturalist.org allows members of the general public to use a combination of computer vision and community-based identifications to document and identify a wide variety of species (iNaturalist 2019a; Van Horn et al. 2017). It falls on the less structured end of the VGI spectrum. Sampling effort is not typically recorded, and the only taxonomic restriction is that observations must be of a living thing. While it is possible for scientists to recruit users to search for specific organisms or organism attributes through the use of iNaturalist projects, the primary purpose of this platform is to document the organisms that the individual user found noteworthy.

While the iNaturalist platform is a tremendous tool for education, the unstructured nature of the observations means that the datasets must be treated with careful consideration. Organism observations on this platform are not a complete record of where they occur, but rather a record of where an interested human was present, observed the organism, and decided to record that observation. These records can be a biased or non-representative spatial sample of the full domain occupied by organisms (Mueller et al. 2019). For the most part, observations reflect only the presence of an organism (not the absence), although a recent feature of iNaturalist allows users to document sampling effort and absence of target organisms (iNaturalist 2019b). These limitations are common to all unstructured and opportunistic datasets collected on VGI platforms (Van Strien et al. 2013).

Despite the limitations with iNaturalist observations, this platform and its associated dataset show potential for supporting vegetation and land cover mapping (Fonte et al. 2015). The species-specific nature of iNaturalist observations seem particularly useful as reference data for vegetation community mapping, although typically the scale of an individual observation would not be a good match for mapping a vegetation community. The presence of a single individual might have little influence on the overall vegetation community, which is defined by percent cover thresholds of diagnostic species. These thresholds are defined by each mapping project, although examples of the mapping rules that define each category are maintained in California by the Manual of California Vegetation (Sawyer et al. 2009).

In a recent mapping project, a subset of the authors of this article used a semi-automated, object-based image analysis approach to map vegetation community types in an open-canopy shrubland (Uyeda et al. 2019). We first mapped discernable shrub species and non-shrub growth-form and land-cover types at the scale of individual shrubs and small patches, then applied percent cover thresholds at coarser scales to map vegetation assemblages at the level of the community. This approach is similar to how human image interpreters perform mapping, but due to the automated approach, it can be efficiently applied to large extents. We assessed the accuracy of the community-level map, as this was the level at which the map was intended to be used. It was not possible to conduct a full accuracy assessment of the fine-scale map due to budget limitations, but identifying errors in this intermediate mapping product could have helped reduce errors propagated forward into the community-level map.

Opportunistic iNaturalist observations of the shrub taxa mapped in the initial species-level mapping effort represent a potentially valuable accuracy assessment dataset. While structured VGI platforms with trained volunteers and specific research goals have been shown to be effective for mapping a limited number of non-native plant species (Hawthorne et al. 2015; Wallace et al. 2016), similar success has not been documented for unstructured VGI platforms and a diverse mix of plant species. The objective of this study is to assess the utility of iNaturalist observations as a reference dataset for species-level vegetation mapping. We address the following research questions: (1) How do the locations of species recorded using iNaturalist compare to where those species have been mapped and (2) What are possible explanations for areas of disagreement?

Methods

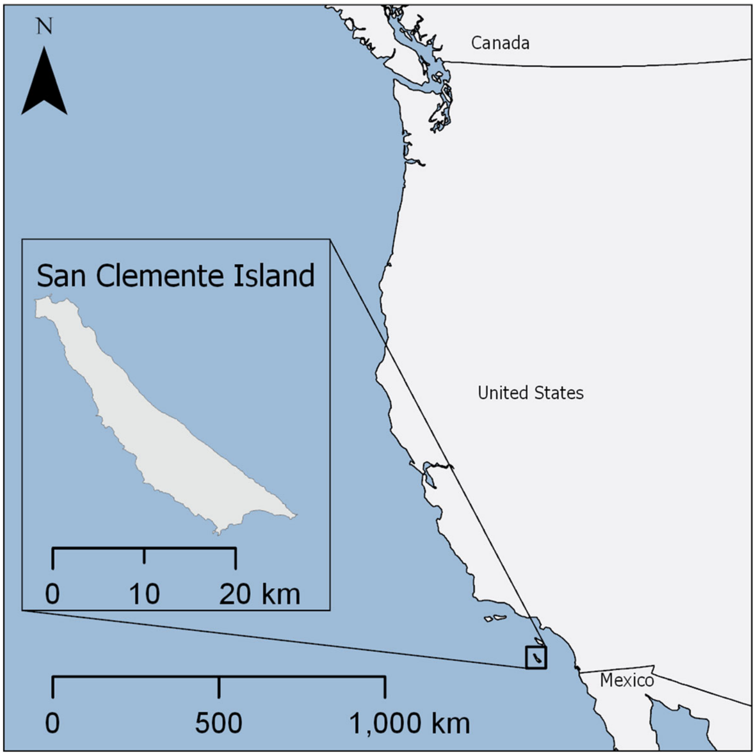
Study area

The study area is San Clemente Island (SCI), the southernmost of the California Channel Islands (Fig. 1). It has been owned and operated by the US Navy since 1934, and has historically been subject to heavy grazing by exotic herbivores. Sheep, cattle, mule deer, and goats have all been present at some point on SCI, with goats being the last to be removed in the early 1990s (Schoenherr et al. 2003). The most abundant vegetation communities are grasslands, maritime succulent scrub alliances dominated by *Opuntia littoralis* or *Lycium californicum*, and coastal sage scrub dominated by *Artemisia californica* (Uyeda et al. 2019). The trees *Lyonothamnus floribundus* ssp. *aspleniifolius*, *Prunus ilicifolia* ssp. *lyonii*, *Quercus tomentella*, and the chaparral shrub *Rhus integrifolia* are often found in canyons. Due to the history of grazing, the recovery of woody perennial species has been regarded with particular importance (Tierra Data Inc. 2011).

Overall approach

We measured the distance from each iNaturalist observation to the closest place it could be verified using two data sources. One source was the fine-scale map product developed to support community-level mapping, and the second was based on manual image interpretation of high-spatial resolution (4–15 cm) aerial imagery. Comparing the presence within the map allowed for the detection of potential undermapping, while comparing visual image interpretation helped to narrow down the source of the undermapping error. If no mapped polygons of a given taxon were found in close proximity to an iNaturalist observation, but the aerial imagery

Fig. 1 General location of San Clemente Island



revealed individuals of that taxon nearby, it would indicate that the mapping error is responsible for the discrepancy. However, if both the mapping and aerial imagery indicate the absence of a taxon in an area, it points instead to a possible error in the positional accuracy of the iNaturalist observation, or perhaps that it was simply not possible to detect the individual using aerial imagery.

iNaturalist observations

Observations of the ten target taxa (*Artemisia* spp., *Baccharis pilularis*, *Rhus integrifolia*, *Cylindropuntia prolifera*, *Opuntia* spp., *Lyonothamnus floribundus* ssp. *aspleniifolius*, *Prunus ilicifolia* ssp. *lyonii*, *Quercus tomentella*, *Lycium californicum*, and non-native Aizoaceae) were exported from iNaturalist on 4 Sept 2019. Each observation contains at least one photograph, the taxon name, coordinates and error associated with those coordinates, time and date that the photograph was taken, the login name of the observer, and any other metadata recorded by the observer. The observation time, date, coordinates, and associated positional error are automatically detected (if available) from the camera metadata found within the digital photograph. Currently, the photos recorded on mobile devices running the iOS operating system (Apple Inc.) have positional error metadata automatically recorded (Apple Inc. 2020). Users of Android mobile devices can automatically record positional error values by taking photos within the iNaturalist mobile application, or estimates of positional error can be manually added to the observation. Observations without positional error were assumed to have values of 10 m or less. A recent study of Android device accuracy found root mean square error of positional accuracy ranging from 2 to 11 m. Error at the low end of this range was found in open canopy conditions, and closed canopy conditions at the high end (Tomaštík et al. 2017). Considering that SCI is dominated by open shrublands, we consider an estimate of 10 m for observations with unspecified positional error to be reasonable, but it is possible that the true value is higher for some of these observations.

Although it is possible on iNaturalist to filter by "Research Grade" status (observations identified to the species level with the consensus of two-thirds of identifiers), the status does not necessarily mean that the identification is correct, as even knowledgeable users sometimes make mistakes (Austen et al. 2018). Therefore, all observations were scrutinized for curation of the data set used here. Filtering by Research Grade status was also not necessary because we only needed to identify *Artemisia* and *Opuntia* to the genus level. The functionally and visually similar species found on the island within each genus (*A. californica*, *A. nesiotica*, *O. littoralis*, *O. oricola*, *O. ficus-indica*) were not separated during the initial vegetation mapping stage, and we maximized the number of records by including all species from each genus. The non-native Aizoaceae category included the four species that were initially targeted for mapping: *Mesembryanthemum crystallinum*, *Mesembryanthemum nodiflorum*, *Carpobrotus edulis*, and *Carpobrotus chilensis*.

Observations with a reported positional error of greater than 10 m were discarded. Observations collected by author KAU were also discarded, as were those in close proximity to species records provided by SCI botanists. These records had already been used to support the creation of the original map. The numerous observations by author CHR were retained, as these primarily consist of records made to document the plants from which terrestrial invertebrates were surveyed as part of an independent research project. Observations of a given species within 10 m of one another were removed, while only the farthest spaced observations in a cluster were retained to maximize distance between observations. In order to reduce possible bias in observation selection, the vegetation map, imagery, and observation photos were not viewed while removing closely spaced points. Clusters of observations within 10 m of one another were retained if they were of different species. Observations in highly developed areas were also removed. Developed areas were mapped as a single "developed" class in the original vegetation map, so no species-level mapping information was available in the immediate vicinity of these observations.

We reviewed the photos associated with each observation again to visually estimate whether the individual pictured was large enough to have been mapped. For trees (*L. floribundus* ssp. *aspleniifolius*, *P. ilicifolia* ssp. *lyonii*, *Q. tomentella*), we used a height threshold of 2 m, a 50 cm threshold for woody shrubs (*Artemisia* spp., *B. pilularis*, *R. integrifolia*), 20 cm for cacti (*C. prolifera*, *Opuntia* spp.) and *L. californicum*, and an area-based threshold of 1 m^2 for non-native Aizoaceae. Although the height or area was not obvious in some photos, this allowed us to remove observations that clearly depicted very small individuals. We also noted when observations were clearly located underneath the canopy of a large tree.

Image properties

Imagery was flown in November 2015, August 2017, and November 2018 with a dual camera system consisting of two Canon Mark II 5D Digital SLR cameras (21 megapixel), with one camera capturing imagery in the visible spectrum and one modified to capture near infrared (NIR) imagery. The sensor system was operated on a Cessna 182 fixed-wing aircraft. Coverage varied with the purpose of the imaging mission, and includes full-island coverage in 2015 and large subsets in 2018, and intermittent image frames were captured along the flight lines in 2017 and 2018. The 2015 imagery was the primary imagery source for the original vegetation mapping and was georeferenced as described in Uyeda et al. 2019. Image frames were georeferenced individually using the 2015 orthoimage mosaic as a reference for geometric control. Acquisition dates, spatial resolution, and coverage of all imagery sources are given in Table 1.

Map properties

The fine-scale vegetation map was created by semi-automated image classification as input for a final community-level map. The focus of this map was to identify shrub and tree species that served as diagnostic species for the community-level map and were clearly visible in the imagery, although other growth forms such as grass and bare ground were also mapped. The primary data sources for this intermediate map were the 2015 aerial imagery and a canopy height layer based on lidar (light detection and ranging) data acquired in 2014. Additional datasets used include herbarium records, previous field surveys, and mapping field work. An object-based image classification with a rule-based expert system was applied using the software program eCognition Developer version 9. In an object-based approach, spatially contiguous pixels are first grouped into segments, or "objects," and classification is applied at the object level. A rule-based system was then used by trained image interpreters to classify segments based on spectral, height, and contextual information. Classification rules were developed over small areas, then tested over larger areas to ensure consistent results. Although the initial goal was to classify the entire island using a single ruleset, it was necessary to make adjustments to the rules in different regions of SCI to increase mapping accuracy. The minimum mapping unit of the fine-scale map is 0.25 m^2. This map was manually edited by trained image interpreters before it was used as input for the community-level map. The community-level map was produced by assessing the cover of key species or cover types over a minimum mapping unit of 0.25 ha, with mapping categories and percent cover thresholds based on the Manual of California Vegetation (Sawyer et al. 2009). Individual trees and chaparral shrubs were also mapped as part of this mapping effort.

Table 1 Imagery products used to detect matching taxa for each observation

Date	Spatial resolution	Coverage
November 2018	0.04–0.06 m	Individual frames
November 2018	0.13 m	5 × 3 km coverage
August 2017	0.04–0.06 m	Individual frames
November 2015	0.15 m	Full-island coverage

The accuracy of the community-level map was assessed using ultra-high spatial resolution imagery (4–6 cm) and 531 accuracy assessment polygons with an area of 0.25 ha. A trained image interpreter independent of the mapping effort reviewed the imagery and estimated percent cover of each distinguishable species or cover type, and 46 accuracy assessment polygons were also reviewed in the field. Accuracy of the tree and chaparral shrub portion of the map was assessed by comparing mapped individuals with independent records of trees and shrubs across SCI. Full details of the mapping approach used are given in Uyeda et al. 2019.

Comparison of presence

The distance from each iNaturalist observation to the edge of the closest mapped individual patch of the corresponding species was recorded as a whole number value using ArcMap 10.7.1 (ESRI Inc). A trained image interpreter performed the same task using the highest-resolution imagery available for the location of each observation. The canopy height layer used in the original mapping effort was occasionally used in image interpretation. However, the canopy height did not rule out the presence of a taxon, as this dataset has been documented to underestimate height in areas with very dense shrub cover (Snavely et al. 2019). An example of the datasets used is shown in Fig. 2. The non-native Aizoaceae (*Mesembryanthemum nodiflorum*) shown in part b of this figure is an example of an individual below the specified area-based threshold.

Results

A total of 526 observations of the ten taxa were initially downloaded, of which 333 had positional error values

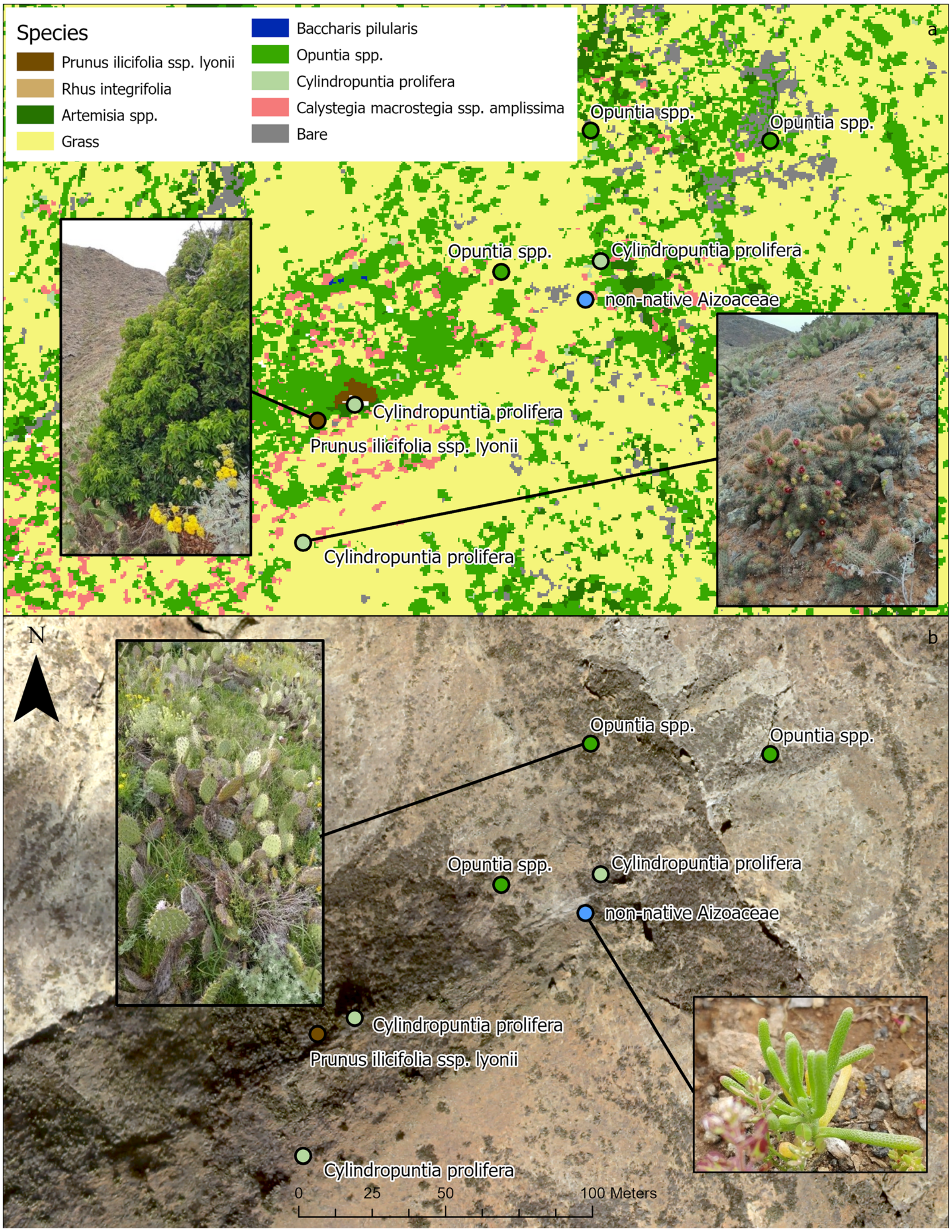


Fig. 2 **a** Example of classified vegetation map and iNaturalist observations with a subset of associated photographs **b** High-resolution imagery and iNaturalist observations. All photographs from iNaturalist.org by Casey Richart

that did not exceed the 10-m threshold established for this study. The discarded observations had positional error values ranging from 11 to 20,327 m, with an average of 1546 m. Removal of observations within 10 m of the same species left a total of 223 iNaturalist observations distributed across the extent of SCI (Fig. 3). Positional error was not recorded (and assumed to be 10 m or less, as described in the methods) in 117, or 52% of these observations. All observations were collected between April 2016 and Aug 2019, with 89 unique observation dates. A total of 16 observers contributed observations, although observations were concentrated within four particularly prolific users who contributed 77% of the total observations. The majority of observations ($n = 195$, 87%) were within 10 m of an individual of the corresponding species based on the imagery (Fig. 4). The number of observations within 10 m of the appropriate species based on the vegetation map is somewhat lower ($n = 143$, 64%). An interactive web map showing the original fine-scale vegetation map, as well as the iNaturalist observations, associated metadata, and measured distances can be found at: https://arcg.is/1GLPH0.

The species-level results are similar to the overall results, with average distance from iNaturalist observation to the presence of the corresponding taxa based on imagery below 10 m for seven of the ten taxa (Table 2). The taxon with the lowest average distance for imagery and mapping comparisons was *Lycium* spp. (2.4 m for imagery, 3.5 m for map). *Cylindropuntia prolifera* was the taxon with the highest average distance (17.8 m) for imagery, and *Q. tomentella* had the highest average distance for the map (245.4 m, Table 2). Maximum values ranged from 12 m for *Artemisia* spp. to 268 for *Opuntia* spp. Both average and maximum differences were greater for the mapping results, with maximum values from 17 to 2122 m (Table 2). Removing observations that were below size-based thresholds or obscured by tree canopy resulting in lower average distances for nearly every taxa, in both the imagery and mapping comparisons (Table 3).

The 2015 imagery (15-cm spatial resolution) was used for the majority of the image comparisons, as was expected considering that full-island coverage exists for this imagery, and the finer-scale imagery sources have smaller coverage areas (Table 1, Table 4). Average distance was similar for all imagery sources.

Discussion

Comparison of iNaturalist and mapping/imagery data sources

Comparison of plant taxa from iNaturalist with taxa attributes represented on a growth form and shrub/tree species level map and derived from visual image interpretation showed close agreement. The agreement with the semi-automated classification map is lower than with direct visual interpretation of imagery, which likely indicates that there is greater error associated with the map. The locations of the trees (*L. floribundus* ssp.

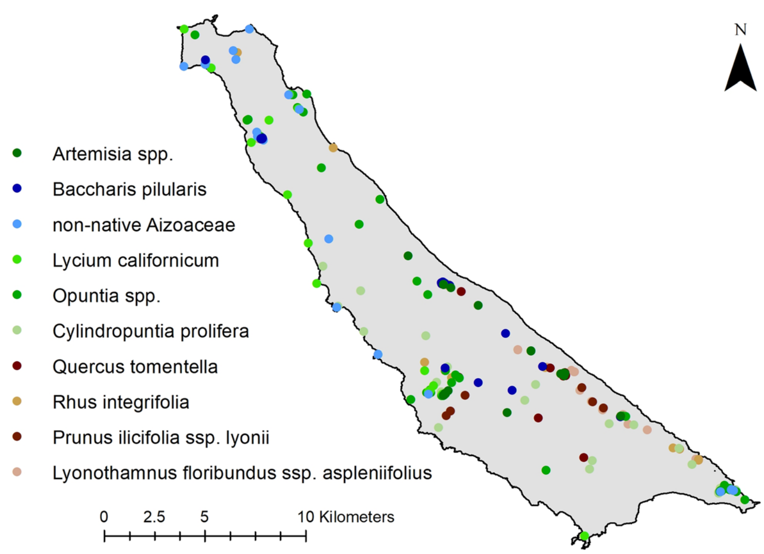


Fig. 3 Map of iNaturalist observations used in this analysis. Some observations were collected in close proximity to one another, so not all 223 individual observations are visible on this map

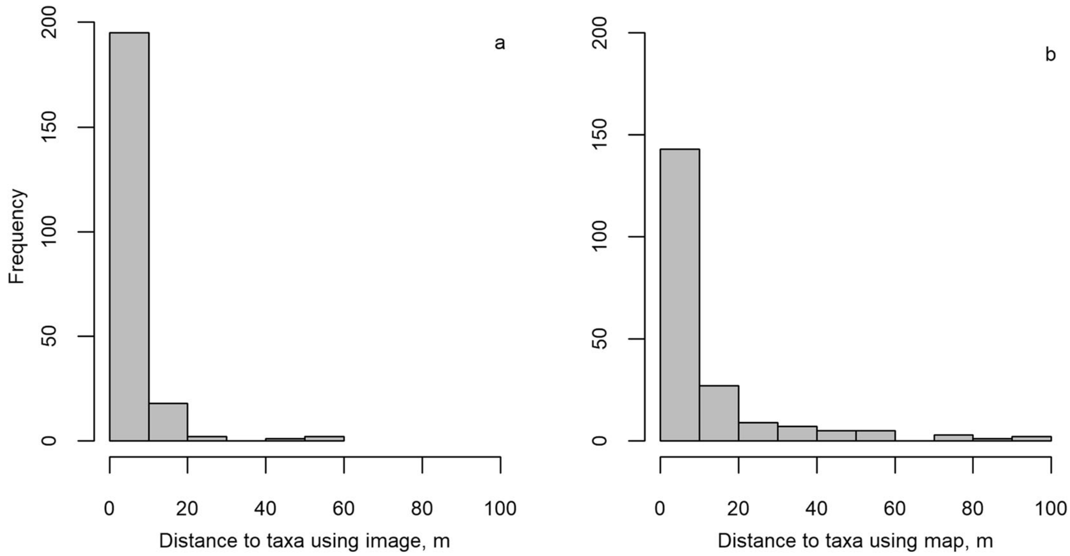


Fig. 4 **a** Histogram of distance from iNaturalist observation to corresponding taxon based on imagery, with only values less than 100 m shown (218 of 223 possible observations shown). **b** Histogram of distance from iNaturalist observation to mapped taxon presence, with only values less than 100 m shown (202 of 223 possible observations shown)

aspleniifolius, *P. ilicifolia* ssp. *lyonii*, *Q. tomentella)* had particularly low agreement with semi-automated mapping results. Low mapping accuracy was documented for trees in the community map, and is attributed to the fact that the canyons in which these trees typically grow are often poorly illuminated in aerial imagery, with dense patches of shade that make species identification challenging (Uyeda et al. 2019). In addition, the restricted distribution of these trees within canyons means that if a tree was not mapped within a particular canyon, the next canyon might be a considerable distance away, as demonstrated by the high maximum distance value for *Q. tomentella* in Table 3. *Cylindropuntia prolifera* also showed low agreement in both the image interpretation and mapping results. This species is often very difficult to detect in the imagery, particularly when it is present as scattered individuals rather than dense aggregations. Low accuracy of the vegetation community associated with this species was also noted in the original mapping study (Uyeda et al. 2019).

Table 2 Distance from iNaturalist observation to closest individual of corresponding taxa based on imagery and map for full set of observations

Taxa	Count	Average distance (imagery), m	Maximum distance (imagery), m	Average distance (mapped), m	Maximum distance (mapped), m
Artemisia spp.	18	4.2	12	6.2	19
Baccharis pilularis	24	3.9	18	44.0	170
Cylindropuntia prolifera	32	17.8	197	35.0	586
Non-native Aizoaceae	22	4.2	13	34.4	235
Lycium californicum	18	2.4	13	3.5	17
Lyonothamnus floribundus ssp. aspleniifolius	15	6.2	18	93.0	702
Opuntia spp.	48	7.8	268	11.7	283
Prunus ilicifolia ssp. *lyonii*	10	15.6	115	32.7	123
Quercus tomentella	11	12.8	52	245.4	2122
Rhus integrifolia	25	9.8	106	18.2	98
Grand total	223	8.4	268	38.3	2122

Table 3 Distance from iNaturalist observation to closest individual of corresponding taxa based on imagery and map for subset of observations above size-based thresholds and not obscured by a tree canopy

Species	Count	Average distance (imagery), m	Maximum distance (imagery), m	Average distance (mapped), m	Maximum distance (mapped), m
Artemisia spp.	13	3.5	12	6.9	19
Baccharis pilularis	18	3.9	18	27.9	170
Cylindropuntia prolifera	26	14.7	197	38.3	586
Non-native Aizoaceae	12	3.3	8	21.0	110
Lycium californicum	18	2.4	13	3.5	17
Lyonothamnus floribundus ssp. *aspleniifolius*	13	6.2	18	47.8	158
Opuntia spp.	46	2.2	13	5.8	78
Prunus ilicifolia ssp. *lyonii*	9	4.6	14	23.6	123
Quercus tomentella	8	10.0	42	319.0	2122
Rhus integrifolia	21	10.8	106	15.7	98
Grand total	184	6.0	197	32.0	2122

Due to the presence-only nature of iNaturalist records, it was not possible to use such data as a spatially sampled reference source for conducting a true accuracy assessment of the species-level map. Observations of the presence of one species do not directly record whether any other species are absent, so it is not possible to use iNaturalist observations to detect over-mapping of some species or growth form classes. For most of the observations, there is no way of knowing if an individual plant represented in the map or identified in the imagery is in fact the same as the one originally photographed. However, the results of this study indicate that iNaturalist observations show great potential to improve vegetation mapping results by revealing areas of undermapping. Ground-based photos are an important reference source for any mapping project, and are particularly valuable for assessing map accuracy when access to field sites is limited.

Table 4 Distance from iNaturalist observation to individual of corresponding taxa by image source

Image date, spatial resolution	Count	Average distance (m)
2015, 15 cm	157	7.2
2017, 5 cm	51	10.5
2018, 13 cm	14	11.6
2018, 5 cm	1	2

Sources of disagreement

It can be tempting to consider field-based photographs as a definitive record of ground conditions. However, it is important to emphasize that iNaturalist observations are reference datasets, and like all reference datasets, have the potential to include error (Jensen 2007). Although iNaturalist observations with high values for positional error were removed from the analysis, over half of the otherwise suitable observations we used for this analysis lack positional error metadata. While we consider it worthwhile to assume that the positional error of these observations is similar to what has been reported for typical modern smartphones GPS accuracy, positional error of nearly 100 m has also been reported (Merry and Bettinger 2019). This positional error might only be noted in the observation metadata if the user noticed the position recorded was incorrect. Interpretation of imagery is also a somewhat subjective process (Powell et al. 2004), and even the results based on the highest resolution imagery assessed here should not be considered to be error-free.

We did not attempt to independently verify which of the data sources was most reliable, although the inclusion of photographs in the iNaturalist observations allowed for the removal of observations that could not be expected to be distinguished based on aerial imagery. Observations of very small individuals are important, and could serve as evidence of recruitment. Rather than thinking of the disagreement caused by these small individuals as an error, we consider it an opportunity to improve the map in the future using field-based VGI input.

Assessment of iNaturalist datasets for vegetation mapping

A substantial strength of the iNaturalist dataset is in the association of field-based photos with every observation. Information that might not have been of interest to the original observer (height, vegetation condition, nearby species, etc.) is nonetheless recorded for many of the photos.

Another obvious benefit of the iNaturalist approach is that there is no direct cost associated with data acquisition, or indirect environmental cost. The latter can be measured both by reduced transportation needs and reduced disturbance of the site. Although both fuel costs and disturbance associated with small field crews is typically minimal, future researchers can help extend the benefit of previous fieldwork by making use of data that has already been collected.

Herbarium records have properties similar to iNaturalist observations and have been commonly used in scientific studies. Records from the Consortium of California Herbaria (CCH2) (http://www.cch2.org) were used as training data for trees and chaparral shrubs in the original map (Uyeda et al. 2019). Herbarium records are commonly used as a source of phenological information such as timing of flowering (Miller-Rushing et al. 2006) or growth of annual plants. While an advantage of these data sources is the long time period of collection (collections from CCH2 start in the early 1800s), the coordinates for older records are often based just on place names.

Collecting herbarium records involves pressing a physical sample of the plant and recording detailed metadata. This greater effort per sample means that fewer records are collected compared to iNaturalist. For example, for SCI, there currently exist 32 records of *R. integrifolia* on CCH2, with only two collected since 2000. There are currently 49 iNaturalist records of *R. integrifolia*, all collected between 2016 and 2019. Herbarium records are an excellent source of information for historical information, but lack the recent, accurate location information of iNaturalist observations. The two data sources are often combined to draw upon the strengths of each (i.e., Biederman et al. 2018; Hereford et al. 2017). Data from iNaturalist, herbaria, and many other databases can be accessed from the Global Biodiversity Information Facility (https://www.gbif.org/), although it is important to note that only "Research grade" iNaturalist observations are included in this database. As described earlier, limiting the query to Research Grade observations is not always desirable.

The restricted access to SCI means that this study area is a little different than other potential sites. The most obvious difference is in the number of observations. Sites with higher rates of visitation tend to have more complete species records (Jacobs and Zipf 2017). However, unlike areas with public access, there is no expectation that researchers will stay on roads and trails. These off-trail observations are valuable both because they were more difficult to obtain and because they could potentially represent less disturbed areas. Another difference is that automatic geo-obscuring enabled in iNaturalist can be turned off for sensitive species due to the restricted access to SCI. Sensitive species are by default obscured by the addition of random noise to their displayed location (although the user who posted the observation is still able to view the true location). It is possible to ask individual users to share the true coordinates of their observations, but the process is laborious. Assessment of whether the default setting of geo-obscuring is necessary for a given species is conducted within the iNaturalist community. Users with "curator" status can weigh in as to whether there is a true conservation risk to sharing a given species' coordinates.

The best dataset for a particular purpose is always going to be the one collected specifically for the research question of interest. We would not recommend that researchers completely abandon their own data collection simply because cheaper data sources are available. iNaturalist is primarily a way for users to document the organisms they find interesting, and researchers must keep the non-random nature of data collection in mind when using the data. Datasets found on iNaturalist are different than what researchers might collect on their own, with advantages and limitations for key data attributes detailed in Table 5. However, with careful consideration of the nature of the data, iNaturalist can be a valuable way for researchers to broaden their reach and make use of a continually growing wealth of information. The use of freely available iNaturalist records is part of a larger trend within the fields of remote sensing and ecology to use open access data to support conservation efforts (Pimm et al. 2015; Rocchini et al. 2017).

Table 5 Comparison of attributes of traditional project specific reference data collection with iNaturalist datasets

	Project specific reference data collection	iNaturalist dataset
Spatial coverage	Expensive to acquire wide coverage, but effort is consistent	Observations are often widely distributed, but effort is likely inconsistent
ID accuracy	Impossible to verify without sending additional field crews	Easily verified using photographs and crowd-sourced identifications
Data type	Quantitative measurement of every plant species present, random sampling	Presence-only data on selected species (no detection of over mapping), possible bias towards species presence in habitat margins or unusual areas
Positional uncertainty	Low positional uncertainty	Positional uncertainty usually unknown, could be as high as 100 m
Applicability to other research projects	Datasets usually not shared, information collected is limited	Datasets shared by default, additional attributes shown in photograph might be useful for others
Labor cost	Significant labor costs	Labor costs already paid through other projects
Environmental cost	Airplane travel, vehicle travel, site disturbance	Environmental costs already incurred through other projects
Summary of benefits	Consistent sampling effort, quantitative measurement, directly addresses project questions	Labor and environmental costs already incurred, data easily verified and shared
Summary of limitations	Expensive, data usually not verified or shared	Inconsistent sampling effort, presence only data, positional uncertainty, data not suitable for all projects

Future applications for vegetation mapping

A valuable application of iNaturalist datasets for vegetation mapping projects in the future could be to use observations as one source of training data for classification using a machine learning approach, expert rule-based classifier, or manual image interpretation. The most reliable training datasets are acquired by visiting field sites that include all classes of interest immediately prior to mapping work. If recent iNaturalist observations show sufficient coverage of some classes of interest in a subset of the project area, any project-specific field effort could be focused on areas or classes not already represented. This would allow for more complete coverage of classes and of the project area. Crowd-sourcing and other forms of VGI have already been used in land cover mapping (Fonte et al. 2015; Waldner et al. 2019), but these are typically done through programs that require recruitment of volunteers and/or highly structured data collection rather than the opportunistically obtaining data that has already been collected.

Location accuracy is an important limitation to the use of iNaturalist observations for training data. In closed-canopy systems with large homogeneous patches of single species, positional error might not be problematic. However, in an open-canopy system with small heterogeneous patches of vegetation (such as SCI), it would likely be necessary to individually correct the position of each observation, possibly introducing error. Even when positional error is low, photographs that show the extent of a large tree or shrub must be taken from several meters away. This discrepancy in the location of the photographer compared to the subject of the photo introduces ambiguity as to the true location of the subject (Keßler et al. 2009). While including the entire plant as well as surrounding vegetation in a photograph is desirable for establishing context within the landscape, observations taken at a distance from the intended organism must have their locations individually assessed to ensure accuracy of the actual organism's location. The widely used imagery from National Agriculture Imagery Program has a horizontal accuracy within 6 m (USDA Farm Service Agency 2020) The positional error in the aerial imagery, combined with the error from a smartphone GPS, makes it unreasonable to expect that iNaturalist observations would be closely co-aligned with the exact shrub that was the original subject of the photograph. Incorrectly classified training data degrades the quality of the final classification, and error can be worse when confusion is between similar classes (Foody et al. 2016).

The SCI vegetation mapping project detailed in Uyeda and others (Uyeda et al. 2019) could have benefited from using iNaturalist observations as a training dataset, but most of the observations were collected after the project was already complete. As iNaturalist records accumulate over the course of years, the database could also serve as an important source of ground reference data for change analysis. It should be noted that we did not attempt to perform an accuracy assessment on the alliance-level map, as at this scale, it is important to know vegetation cover thresholds to properly categorize the landscape. The observations we had available were often focused on only the plant of interest, without any indication of how much total cover of the plant was present in a given area. However, with enough observations that show the surrounding landscape, it could be possible for future alliance-level vegetation maps to incorporate iNaturalist observations into both the training and accuracy assessment datasets, particularly for easily distinguished alliances.

Conclusions

iNaturalist and other sources of opportunistic VGI provide an excellent opportunity for researchers to make use of data that have already been collected. Positional accuracy is an important consideration for vegetation mapping at the scale of the individual plant, particularly in an open-canopy shrubland. We suggest that researchers consider incorporating iNaturalist observations into their research, and contribute iNaturalist observations themselves.

Acknowledgments This research would not have been possible without the contributions of numerous iNaturalist users. In addition to the observations of author CHR, iNaturalist users whose observations were used in this research include cedric_lee, vireolanius, shrike2, dmathews, jrebman, stephanomeria, serpophaga, elizabercel, empid, sulavanderplank, madpurdy,stephlbartlett, sue_meiman, terrilldactyl, demi5, jennyh, sabrinamashburn and taylorsteichen.

Availability of data and material (data transparency) Available at https://arcg.is/1GLPH0

Code availability Not applicable.

Funding information This project was funded under agreement (Cooperative Agreement Award No: W9126G-18-2-0068) by the U.S. Army Corps of Engineers.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

Apple Inc. (2020). Apple Developer Documentation. https://developer.apple.com/documentation/corelocation/clvisit/1614679-horizontalaccuracy#declarations.

Austen, G. E., Bindemann, M., Griffiths, R. A., & Roberts, D. L. (2018). Species identification by conservation practitioners using online images: accuracy and agreement between experts. *PeerJ, 6*, e4157. https://doi.org/10.7717/peerj.4157.

Beaubien, E. G., & Hall-Beyer, M. (2003). Plant phenology in Western Canada: trends and links to the view from space. *Environmental Monitoring and Assessment, 88*(1–3), 419–429. https://doi.org/10.1023/A:1025597829895.

Biederman, L. A., Derek, A., Nancy, S., John, P., Judy, B., & Jeanne, P. (2018). Using phenological monitoring in situ and historical records to determine environmental triggers for emergence and anthesis in the rare orchid Platanthera praeclara Sheviak & Bowles. *Global Ecology and Conservation, 16*, e00461. https://doi.org/10.1016/j.gecco.2018.e00461.

Conrad, C. C., & Hilchey, K. G. (2011). A review of citizen science and community-based environmental monitoring: issues and opportunities. *Environmental Monitoring and Assessment, 176*(1–4), 273–291. https://doi.org/10.1007/s10661-010-1582-5.

Eitzel, M. V., Cappadonna, J. L., Santos-Lang, C., Duerr, R. E., Virapongse, A., West, S. E., et al. (2017). Citizen science terminology matters: exploring key terms. *Citizen Science: Theory and Practice, 2*(1), 1. https://doi.org/10.5334/cstp.96.

ElQadi, M. M., Dorin, A., Dyer, A., Burd, M., Bukovac, Z., & Shrestha, M. (2017). Mapping species distributions with social media geo-tagged images: case studies of bees and flowering plants in Australia. *Ecological Informatics, 39*(February), 23–31. https://doi.org/10.1016/j.ecoinf.2017.02.006.

Elwood, S., Goodchild, M., & Sui, D. (2013). Crowdsourcing geographic knowledge: volunteered geographic information (VGI) in theory and practice. *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice, 9789400745*, 1–396. https://doi.org/10.1007/978-94-007-4587-2.

Fonte, C. C., Bastin, L., See, L., Foody, G., & Lupia, F. (2015). Usability of VGI for validation of land cover maps. *International Journal of Geographical Information Science, 29*(7), 1269–1291. https://doi.org/10.1080/13658816.2015.1018266.

Foody, G. M., Pal, M., Rocchini, D., Garzon-Lopez, C. X., & Bastin, L. (2016). The sensitivity of mapping methods to reference data quality: training supervised image classifications with imperfect reference data. *ISPRS International Journal of Geo-Information, 5*(11). https://doi.org/10.3390/ijgi5110199.

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal, 69*(4), 211–221. https://doi.org/10.1007/s10708-007-9111-y.

Hawthorne, T. L., Elmore, V., Strong, A., Bennett-Martin, P., Finnie, J., Parkman, J., Harris, T., Singh, J., Edwards, L., & Reed, J. (2015). Mapping non-native invasive species and accessibility in an urban forest: a case study of participatory mapping and citizen science in Atlanta, Georgia. *Applied Geography, 56*, 187–198. https://doi.org/10.1016/j.apgeog.2014.10.005.

Heider, K., Lopez, J. M. R., & Scheffran, J. (2018). The potential of volunteered geographic information to investigate peri-urbanization in the conservation zone of Mexico City. *Environmental Monitoring and Assessment, 190*(4), 219. https://doi.org/10.1007/s10661-018-6597-3.

Hereford, J., Schmitt, J., & Ackerly, D. D. (2017). The seasonal climate niche predicts phenology and distribution of an ephemeral annual plant, Mollugo verticillata. *Journal of Ecology, 105*(5), 1323–1334. https://doi.org/10.1111/1365-2745.12739.

iNaturalist. (2019a). What is it. https://www.inaturalist.org/pages/what+is+it. Accessed 24 February 2020.

iNaturalist. (2019b). Trips enable the collection of presence-absence data by adding context to your observations. https://www.inaturalist.org/pages/trips. Accessed 24 February 2020.

Isaac, N. J. B., van Strien, A. J., August, T. A., de Zeeuw, M. P., & Roy, D. B. (2014). Statistics for citizen science: extracting signals of change from noisy ecological data. *Methods in Ecology and Evolution, 5*(10), 1052–1060. https://doi.org/10.1111/2041-210X.12254.

Iwao, K., Nishida, K., Kinoshita, T., & Yamagata, Y. (2006). Validating land cover maps with degree confluence project information. *Geophysical Research Letters, 33*(23), 1–5. https://doi.org/10.1029/2006GL027768.

Jacobs, C., & Zipf, A. (2017). Completeness of citizen science biodiversity data from a volunteered geographic information perspective. *Geo-Spatial Information Science, 20*(1), 3–13. https://doi.org/10.1080/10095020.2017.1288424.

Jensen, J. R. (2007). *Remote sensing of the environment: an earth resource perspective* (2nd ed.). Upper Saddle River: Pearson Prentice Hall.

Keßler, C., Maué, P., Heuer, J. T., & Bartoschek, T. (2009). Bottom-up gazetteers: learning from the implicit semantics of geotags. In International conference on GeoSpatial Sematics (pp. 83–102). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-10436-7_6.

Lonhart, S. I., Jeppesen, R., Beas-Luna, R., Crooks, J. A., & Lorda, J. (2019). Shifts in the distribution and abundance of coastal marine species along the eastern Pacific Ocean during marine heatwaves from 2013 to 2018. *Marine Biodiversity Records, 12*(1), 1–15. https://doi.org/10.1186/s41200-019-0171-8.

Merry, K., & Bettinger, P. (2019). Smartphone GPS accuracy study in an urban environment. *PLoS One, 14*(7), e0219890. https://doi.org/10.1371/journal.pone.0219890.

Miller-Rushing, A. J., Primack, R. B., Primack, D., & Mukunda, S. (2006). Photographs and herbarium specimens as tools to document phenological changes in response to global warming. *American Journal of Botany, 93*(11), 1667–1674. https://doi.org/10.3732/ajb.93.11.1667.

Mueller, M. A., Drake, D., & Allen, M. L. (2019). Using citizen science to inform urban canid management. *Landscape and Urban Planning, 189*(May 2018), 362–371. https://doi.org/10.1016/j.landurbplan.2019.04.023.

Pimm, S. L., Alibhai, S., Bergl, R., Dehgan, A., Giri, C., Jewell, Z., Joppa, L., Kays, R., & Loarie, S. (2015). Emerging technologies to conserve biodiversity. *Trends in Ecology & Evolution, 30*(11), 685–696. https://doi.org/10.1016/j.tree.2015.08.008.

Powell, R. L., Matzke, N., De Souza, C., Clark, M., Numata, I., Hess, L. L., et al. (2004). Sources of error in accuracy assessment of thematic land-cover maps in the Brazilian Amazon. *Remote Sensing of Environment, 90*(2), 221–234. https://doi.org/10.1016/j.rse.2003.12.007.

Richart, C. H., Chichester, L. F., Boyer, B., & Pearce, T. A. (2018). Rediscovery of the southern California endemic American Keeled Slug Anadenulus cockerelli (Hemphill, 1890) after a 68-year hiatus. *Journal of Natural History, 53*, 1515–1531. https://doi.org/10.1080/00222933.2018.1447700.

Rocchini, D., Petras, V., Petrasova, A., Horning, N., Furtkevicova, L., Neteler, M., Leutner, B., & Wegmann, M. (2017). Open data and open source for remote sensing training in ecology. *Ecological Informatics, 40*(May), 57–61. https://doi.org/10.1016/j.ecoinf.2017.05.004.

Sawyer, J. O., Keeler-Wolf, T., & Evens, J. M. (2009). *A manual of California vegetation* (Second ed.). Sacramento: California Native Plant Society.

Schoenherr, A. A., Feldmeth, C. R., & Emerson, M. J. (2003). *Natural history of the islands of California*. Berkeley: Univ of California Press.

See, L., Mooney, P., Foody, G., Bastin, L., Comber, A., Estima, J., Fritz, S., Kerle, N., Jiang, B., Laakso, M., Liu, H. Y., Milčinski, G., Nikšič, M., Painho, M., Pődör, A., Olteanu-Raimond, A. M., & Rutzinger, M. (2016). Crowdsourcing, citizen science or volunteered geographic information? The current state of crowdsourced geographic information. *ISPRS International Journal of Geo-Information, 5*(5). https://doi.org/10.3390/ijgi5050055.

Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C., & Haklay, M.(Muki). (2017). A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science, 31*(1), 139–167. https://doi.org/10.1080/13658816.2016.1189556.

Snavely, R., Uyeda, K., Stow, D., O'Leary, J., & Lambert, J. (2019). Mapping vegetation community types in a highly-disturbed landscape: integrating hierarchical object-based image analysis with lidar-derived canopy height data. *International Journal of Remote Sensing, 40*(11), 4384–4400. https://doi.org/10.1080/01431161.2018.1562588.

Taylor, A. T., Von Schmeling, H., & Long, J. M. (2018). Photographs of wading bird depredation update invasion extent of Monopterus albus (Asian swamp eel). *Southeastern Naturalist, 17*(4), 72–76. https://doi.org/10.1656/058.017.0408.

Tierra Data Inc. (2011). *San Clemente Island vegetation condition and trend analysis, 2010*. San Diego: Southwest Division U.S. Naval Facilities Engineering Command.

Tomaštík Jr., J., Tomaštík Sr., J., Saloň, Š., & Piroh, R. (2017). Horizontal accuracy and applicability of smartphone GNSS positioning in forests. *Forestry, 90*(2), 187–198. https://doi.org/10.1093/forestry/cpw031.

USDA Farm Service Agency. (2020). NAIP Imagery. https://www.fsa.usda.gov/programs-and-services/aerial-photography/imagery-programs/naip-imagery/. Accessed 24 February 2020.

Uyeda, K. A., Warkentin, K. K., Stow, D. A., O'Leary, J. F., Snavely, R., Lambert, J., et al. (2019). Vegetation mapping using hierarchical object-based image analysis applied to aerial imagery and lidar data. *Applied Vegetation Science, 23*(1), 80-93. https://doi.org/10.1111/avsc.12467.

Van Horn, G., Mac Aodha, O., Song, Y., Cui, Y., Sun, C., Shepard, A., et al. (2017). The iNaturalist species classification and detection dataset. In *CVPR* (pp. 4–6).

Van Strien, A. J., Van Swaay, C. A. M., & Termaat, T. (2013). Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. *Journal of Applied Ecology, 50*(6), 1450–1458. https://doi.org/10.1111/1365-2664.12158.

Vandenlangenberg, S. M., Canfield, J. T., & Magner, J. A. (2003). A regional survey of malformed frogs in Minnesota (USA) (Minnesota malformed frogs). *Environmental Monitoring and Assessment, 82*(1), 45–61. https://doi.org/10.1023/A:1021684723301.

Waldner, F., Schucknecht, A., Lesiv, M., Gallego, J., See, L., Pérez-Hoyos, A., d'Andrimont, R., de Maet, T., Bayas, J. C. L., Fritz, S., Leo, O., Kerdiles, H., Díez, M., van Tricht, K., Gilliams, S., Shelestov, A., Lavreniuk, M., Simões, M., Ferraz, R., Bellón, B., Bégué, A., Hazeu, G., Stonacek, V., Kolomaznik, J., Misurec, J., Verón, S. R., de Abelleyra, D., Plotnikov, D., Mingyong, L., Singha, M., Patil, P., Zhang, M., & Defourny, P. (2019). Conflation of expert and crowd reference data to validate global binary thematic maps. *Remote Sensing of Environment, 221*(October 2018), 235–246. https://doi.org/10.1016/j.rse.2018.10.039.

Wallace, C. S. A., Walker, J. J., Skirvin, S. M., Patrick-Birdwell, C., Weltzin, J. F., & Raichle, H. (2016). Mapping presence and predicting phenological status of invasive buffelgrass in Southern Arizona using MODIS, climate and citizen science observation data. *Remote Sensing, 8*(7). https://doi.org/10.3390/rs8070524.